

Universidade de Lisboa

Faculdade de Farmácia



**Quantificação de componentes em formas
farmacêuticas sólidas via espectroscopia
NIR e Resolução Multivariada de Curvas**

Yasmin Rocha de Oliveira

Mestrado Integrado em Ciências Farmacêuticas

2017

Universidade de Lisboa
Faculdade de Farmácia



**Quantificação de componentes em formas
farmacêuticas sólidas via espectroscopia
NIR e Resolução Multivariada de Curvas**

Yasmin Rocha de Oliveira

**Monografia de Mestrado Integrado em Ciências Farmacêuticas
apresentada à Universidade de Lisboa através da Faculdade de Farmácia**

Orientador: Prof. Dr. João Almeida Lopes

2017

Resumo

Este trabalho tem como objetivo desenvolver um protocolo de quantificação simultânea de substância ativa e excipientes, em formulações sólidas produzidas em laboratório, usando espectroscopia de infravermelho próximo e a Resolução Multivariada de Curvas (MCR, do inglês *Multivariate Curve Resolution*). As amostras são misturas de pós, constituídas por paracetamol, como substância ativa e celulose microcristalina, talco e estearato de magnésio, como excipientes. Foi usado um conjunto de 40 amostras, para a calibração dos modelos e um conjunto distinto, com 29 amostras, como teste. Ambos foram produzidos em laboratório, de acordo com um delineamento experimental. As medições foram feitas em modo de refletância num espectrofotômetro com Transformada de *Fourier*.

Foi explorada a MCR, como método de calibração multivariada e os resultados comparados aos do método Regressão Parcial de Mínimos Quadrados (PLS, do inglês, *Partial Least Squares*), de uso corrente nas análises quimiométricas quantitativas. Para esse efeito, construíram-se dois modelos com MCR – um modelo rígido, contendo os espectros puros de cada componente e um modelo flexível. Foi ainda construído um modelo com PLS. Os resultados foram avaliados em termos de erro de calibração, erro de predição e correlação entre as concentrações estimadas e as concentrações reais.

Este método é adequado para aferir a concentração de componentes presentes em maior quantidade na formulação, mas não para os componentes minoritários. Concluiu-se também que a MCR é adequada para análises qualitativas, sendo capaz de reconstituir os espectros puros dos componentes.

Professor Doutor João Almeida Lopes

Palavras-chave: espectroscopia; infravermelho próximo; quimiometria: calibração multivariada; Resolução de Curvas Multivariadas

Abstract

This work aims to develop a Near-infrared based protocol, using the chemometric method Multivariate Curve Resolution (MCR), to access the active pharmaceutical ingredient (API) and excipient concentrations at the same time.

The samples were based on a paracetamol formulation with three excipients: microcrystalline cellulose, talc and magnesium stearate. A set of 40 powders was used to calibrate the models and a set of 29 different powders was projected onto the models. Both calibration and test set were laboratory-scale samples, produced according to an experimental design. The samples were measured in reflectance mode in a Fourier-transform NIR spectrometer.

MCR was used as the multivariate calibration method and its results were compared to the ones given by Partial Least Squares Regression (PLS), which is widely used in quantitative chemometric analysis. With that purpose, two MCR models were built: a hard-model, with pure spectra equality constraint and a soft-model, lacking this additional information. It was also built a PLS model. Results were accessed by means of calibration error, prediction error and correlation between estimated and real data. With this method, is suitable for estimation of major components concentrations, but, not of the minor components. It was also concluded that MCR can estimate the pure spectra of components, and hence, perform qualitative analysis.

Keywords: spectroscopy; near infrared; chemometrics; multivariate calibration.

Agradecimentos

À minha Mãe,
ao meu Professor,
e a Mim.

Abreviaturas

MCR – *Multivariate Curve Resolution*

PLS – *Parcial Least Squares*

NIR – *Near Infrared*

RMSECV – *Root Mean Square Error of Cross Validation*

RMSEP – *Root Mean Square Error of Prediction*

Índice:

1	Introdução	8
1.1	Objetivos	8
1.2	Fundamento teórico	8
1.3	Teoria da MCR	10
1.4	Aspetos da modelagem	12
1.5	Pré-processamento dos espectros	12
1.6	Erro	13
2	Materiais e Métodos	14
2.1	Amostras	14
2.2	Obtenção dos espectros NIR	15
2.3	Modelagem	15
2.4	Tratamento dos resultados	15
2.5	Pré-processamento de dados	16
3	Resultados e discussão	17
3.1	Modelos MCR	17
3.1.1	Modelo A	17
3.1.1.1	Calibração	17
	<i>Loadings</i>	17
	<i>Scores</i>	2
3.1.1.2	Previsão	2
3.2	Modelo B	2
3.2.1	Calibração	2
3.2.2	Previsão	2
3.3	Modelo PLS	2
3.3.1	Calibração	2
3.3.2	Previsão	3
3.4	Comparação dos métodos	3
4	Conclusões	3
	Referências Bibliográficas	4

Índice de Figuras:

Figura 1	11
Figura 2	11
Figura 3	18
Figura 4	20
Figura 5	22
Figura 6	24

Índice de Tabelas:

Figura 1	14
Figura 2	17
Figura 3	19
Figura 4	21
Figura 5	25

1 Introdução

1.1 Objetivos

Este trabalho tem como objetivo geral desenvolver um protocolo de quantificação simultânea de substância ativa e excipientes em formas sólidas, utilizando espectroscopia NIR e Resolução Multivariada de Curvas. Os objetivos específicos são avaliar o desempenho do método MCR, por meio da comparação dos resultados com o método PLS.

1.2 Fundamento teórico

A espectroscopia de infravermelho próximo (NIR, do inglês, *Near Infrared*) tem conquistado uma grande aceitação na indústria farmacêutica, sendo uma ferramenta útil no teste a matérias primas, no controlo de qualidade dos produtos e monitorização dos processos. O interesse crescente por esta tecnologia advém certamente das vantagens perante as técnicas analíticas tradicionais: o facto de ser não invasiva, não destrutiva, a rapidez das medições, e a fácil preparação das amostras, sem necessidade de pré-tratamento, são algumas das que se podem contar(1)(2)(3).

Se é verdade que foi Karl Norris, que na década de 60, reconheceu o potencial da técnica para a indústria, nos dias de hoje é um movimento concertado entre o crescente número de requerimentos por parte dos fabricantes, recentes diretrizes emitidas pelas entidades regulamentares e atenção por parte de investigadores da área, que impulsiona esta tecnologia para a implementação nas rotinas da indústria farmacêutica.

A iniciativa de Tecnologias Analíticas de Processo (PAT initiative, do inglês *Process Analytical Technology*) em 2003 abriu perspectivas para a implementação de tecnologias inovadoras no controle da qualidade e dos processos de produção(4). No sentido de colmatar alguns dos entraves colocados pelos fabricantes e dar impulso a esta iniciativa, a Food and Drug Administration (FDA), publicou um guia endereçado à indústria farmacêutica, onde descreve uma estratégia estruturada dentro dos limites regulatórios e questões técnicas e científicas que acompanham a adoção destas tecnologias inovadoras.(5)

A indústria farmacêutica caminha para um paradigma que preconiza que a qualidade seja construída ao longo do processo (*Quality by Design*), e que seja baseada num profundo conhecimento dos processos(6). Protagonistas desta cruzada são consensuais em afirmar que “a habilidade de predição reflete um bom grau de conhecimento do processo.”(5)

A espectroscopia NIR revela-se perfeitamente enquadrada nesta perspectiva: a independência dos seus componentes instrumentais – sonda, computador, etc – fala pela versatilidade no projetar das medições (permite diversificar os pontos de controle, a posição da sonda, e a medição *in-line*); a rapidez de varredura viabiliza a monitorização em tempo real dos parâmetros críticos de ambos, processo e produtos; a medição *in-line* acarreta a vantagem de diminuir o erro humano(2)(7) Tudo isto contribui para o conhecimento do processo, que por sua vez contribui para a melhoria na gestão do processo, com vista à redução de custos e aumento da produtividade. Sem sombra de dúvida, estas vantagens repercutem-se no crescimento do setor e em ganhos para a saúde pública.

O infravermelho próximo é a zona do espectro de luz que compreende o intervalo de comprimentos de onda entre os 750 e os 2500 nanómetros. Devido aos movimentos vibracionais de combinação e de sobretons dos grupos –OH, -CH, -NH e -SH, detetados nesta zona, é possível a determinação qualitativa e quantitativa das moléculas presentes numa amostra analisada(1)(7).

Uma lacuna do método é a não seletividade das medições efetuadas nesta região do espectro. Este fato, aliado à complexidade física e química das formulações farmacêuticas, gera espectros bastante complexos e difíceis de relacionar às propriedades das amostras. (8)

O processamento do sinal analítico e a extração da informação relevante, é feita com recurso a modelos de regressão matemática multivariados. Dentre estes, o método de Regressão de Quadrados Mínimos Parciais (PLS, do inglês, *Partial Least Squares*), já amplamente utilizado em quimiometria (4) e a Resolução Multivariada de Curvas (MCR, do inglês, *Multivariate Curve Resolution*), discutido neste trabalho. (9)

1.3 Teoria da MCR

As análises químicas de produtos farmacêuticos envolvem frequentemente a medição de sinais de caráter multivariado. A decomposição destes sinais, necessária para a sua interpretação, origina frequentemente dados que se organizam sob a forma de matrizes de segunda ordem (10). A MCR propõe-se a fazer uma análise quimiométrica, conjugando a estatística multivariada com a imposição de restrições com significado físico. As figuras (1) e (2) e a explicação que se segue pretendem elucidar o funcionamento do método.

O primeiro passo deste método é desenvolver a matriz X . Nas linhas constam os espectros das várias amostras e nas colunas os valores de $\text{Log } 1/R$ medido em cada comprimento de onda. Segue-se a estimativa do possível número de componentes, utilizando métodos como Análise de Componentes Principais. É genericamente assumido o funcionamento da lei de Lambert-Beer para as medições espectroscópicas. Assente neste pressuposto, o MCR decompõe uma matriz X , segundo um modelo bi-linear descrito pela **Eq.(1)**:

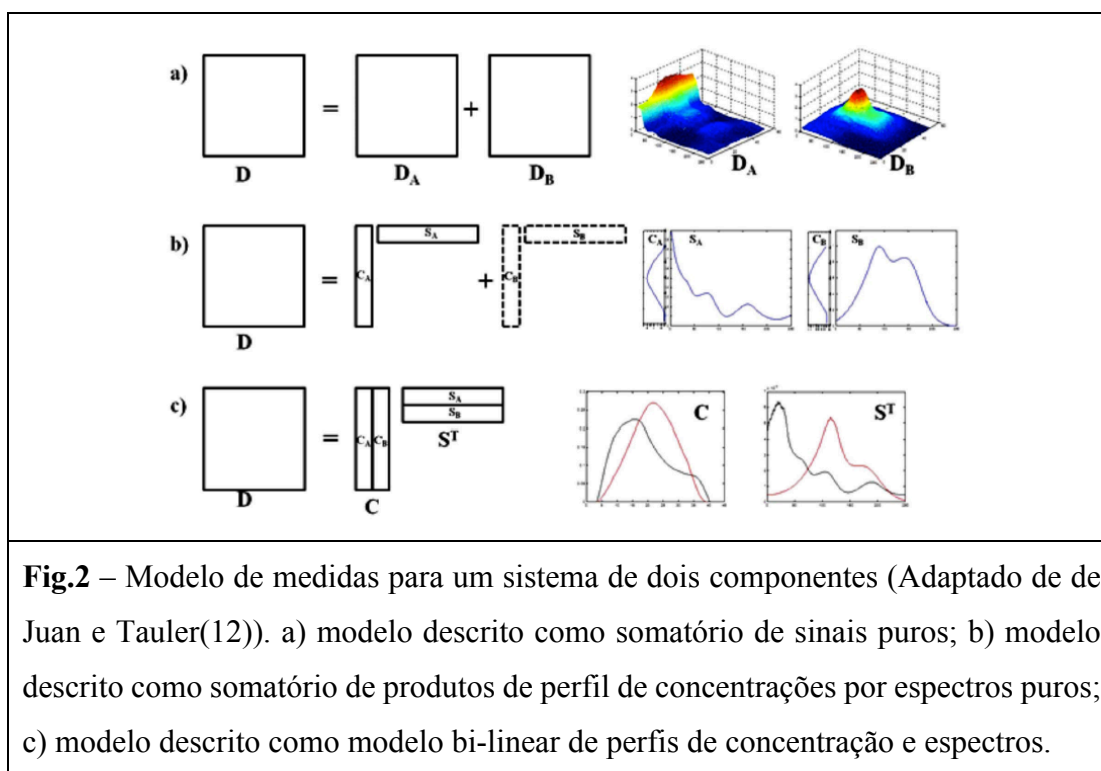
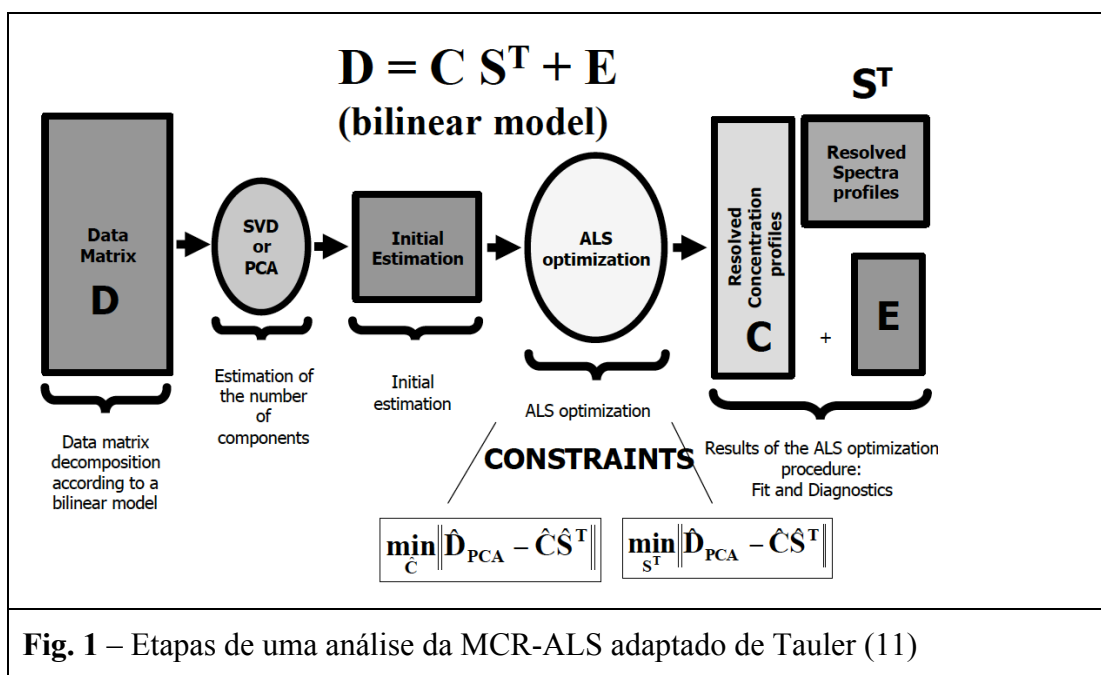
$$X = CS^T + E \quad (1)$$

onde X (I,J) é a matriz de dados experimentais, de dimensão I amostras por J comprimentos de onda; C (I,K) é a matriz de perfis de concentração dos K analitos presentes nas amostras; S^T (K,J) é a matriz de espectros, donde K linhas contêm os espectros puros associados aos K analitos; E (I,J) é a matriz associada aos resíduos (13).

A resolução iterativa das variáveis de C e S é realizada com algoritmo de Mínimos Quadrados Alternantes (ALS, do inglês *Alternating Least Squares*) e requer uma primeira estimativa para C e S^T . A estimativa inicial é executada através de análise exploratória de dados, utilizando métodos como Análise de Fatores Evolutivos (EFA, do inglês *Evolving Factor Analysis*), ou aproximação das variáveis puras(14).

A otimização iterativa decorre, de modo convergente entre a matriz reconstruída X_{est} a partir do cálculo de C e de S e a matriz real X , tal que o erro E seja mínimo. O erro E é descrito pela **Eq.(2)** e é uma medida de exatidão do sistema:(10)

$$E = (X - CS^T)^2 \quad (2)$$



A resolução da Eq.(1) permite soluções matemáticas, sem significado do ponto de vista físico-químico. Isto acontece por meio de um fenómeno designado ambiguidade rotacional. A MCR permite a introdução de informações químicas ou matemáticas no sistema através da imposição de restrições. Dentre as mais usadas existem as restrições

de não negatividade, de aproximação à unidade (*closure*) de unimodalidade, aplicadas em modelos flexíveis, ou modelos rígidos.(15)

1.4 Aspetos da modelagem

Com MCR desenvolveram-se dois modelos: um em que a única informação dada ao sistema foi a matriz de espectros das 40 amostras, e outro em que se introduziu também a matriz de espectros puros. Para o modelo flexível, obtiveram-se e corrigiram-se loadings e scores, enquanto que no modelo rígido apenas scores. As restrições aplicadas em ambos os modelos MCR foram: de não negatividade, tanto para as contribuições C, como para as S e de *closure*, incluindo os quatro componentes. No segundo modelo, foi aplicada a restrição de igualdade para os espectros puros. O número de componentes é um dos passos do programa de software e foi selecionada a linha de resultados correspondente a quatro.

A modelagem por PLS, contempla dois pontos chave: a validação e o pré-processamento de dados. Foram realizados dois tipos de validação. A técnica de validação interna foi feita com base na validação cruzada dos dados, segundo o método designado por *leave-one-out*. Este consiste na exclusão, à vez, de cada uma das 40 amostras do conjunto de calibração, considerando-a como amostra de teste, face a um modelo calibrado com as restantes 39 amostras. A validação externa foi efetuada recorrendo ao conjunto de amostras teste.

1.5 Pré-processamento dos espectros

Os espectros de infravermelho próximo contêm tanto informação química, como informação física. Esta etapa visa remover informações físicas e instrumentais, que não estão relacionadas à informação química de interesse, de modo a melhorar a análise exploratória dos dados e a regressão multivariada subsequentes.(16)

Os métodos que tipicamente se adequam a estes dados são: *Multiplicative Scatter Correction* (MSC) e *Standard Normal Variate* (SNV), representando os métodos de correção de dispersão, projetados para reduzir as diferenças na dispersão da luz, ocasionada pela variabilidade física entre as amostras e ajustar os desvios da linha de base entre as amostras(16); e os métodos de derivação espectral, nomeadamente as

derivadas de Savitsky-Golay. Estes têm a capacidade de remover os efeitos aditivos e multiplicativos entre os espectros. A primeira derivada corrige a linha de base (efeito aditivo) e a segunda derivada ajusta tanto a linha de base como os declives (efeito multiplicativo). Ambas recorrem ao alisamento dos espectros antes do cálculo da derivada, de modo a que a aplicação da derivada não reduza em demasia a razão sinal-ruído(16).

1.6 Erro

O principal indicador de exatidão dos modelos foi o erro quadrático médio da validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*) – Eq.(3):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N_C} (\hat{Y}_{CVi} - Y_i)^2}{N_C}} \quad (3)$$

Na Eq.(3), \hat{Y}_{CVi} , é a estimativa de validação cruzada para a amostra i , Y_i é a concentração real da amostra i e N_C é o número de amostras de calibração.

A validação externa foi avaliada por meio do erro quadrático médio de predição (RMSEP, do inglês, *Root Mean Square Error of Prediction*) – Eq.(4):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_P} (\hat{Y}_{CVi} - Y_i)^2}{N_P}} \quad (4)$$

Na Eq. (4), \hat{Y}_i , é a estimativa de predição para a amostra i , Y_i é a concentração real da amostra i e N_P é o número de amostras do conjunto de teste.

2 Materiais e Métodos

2.1 Amostras

Foram produzidos dois tipos de amostras em laboratório: um conjunto de calibração, constituído por 40 amostras de pós, e um conjunto de teste, constituído por 29 amostras de pós.

Como técnica de amostragem, recorreu-se à aplicação de dois delineamentos experimentais diferentes (*D-optimal mixture design*): um para o conjunto de calibração e outro para o conjunto de teste

A formulação é constituída por quatro componentes: paracetamol, como substância ativa, e celulose microcristalina, talco e estearato de magnésio, como excipientes. Na **Tabela 1**, estão discriminados os intervalos de concentração adotados para cada um dos componentes.

Tabela 1. Intervalo de fração mássica para o conjunto de calibração e de teste

Componentes	Fração mássica de referência	Amostras de calibração		Amostras de teste	
		min	max	Min	Max
Paracetamol	0.83	0.40	0.99	0.58	0.92
Celulose microcristalina	0.1	0.00	0.45	0.00	0.37
Talco	0.05	0.00	0.23	0.00	0.10
Estearato de magnésio	0.02	0.00	0.10	0.00	0.01
Total de amostras		40		29	

A preparação das amostras em laboratório compreendeu, a pesagem de cada um dos componentes, seguida da mistura homogénea dos quatro em almofariz de vidro. Foram obtidos os espectros NIR dos componentes ainda puros. A massa total de cada amostra foi de 3g, com a concentração de sustância ativa a variar entre 40 e 99%. Este procedimento foi seguido tanto para o conjunto de calibração, como para o conjunto de teste.

2.2 Obtenção dos espectros NIR

Os espectros foram recolhidos num espectrofotómetro de infravermelho próximo com transformada de Fourier (Antaris I, ThermoNicolet, Madison, USA), equipado com uma sonda de reflectância de fibra ótica (SabIR, ThermoNicolet), e um detetor de índio-gálio-arsénio. (InGaAs). O equipamento foi controlado através do programa de software Result (ThermoNicolet Industrial Solutions, Madison, USA), próprio para aquisição automática dos espectros. Os espectros de cada amostra foram registados com uma resolução de 2 cm^{-1} e uma média de 64 *scans* num intervalo de número de onda entre 4250 e 10 000 cm^{-1} . As medições foram feitas, recorrendo a uma sonda de fibra ótica, segundo o método de reflectância difusa. Para execução do “branco” – espectro de fundo – foi inserida a sonda de fibra ótica no compartimento de amostragem, contendo a referência (Spectralon). Os espectros dos pós foram obtidos, inserindo a sonda diretamente nas amostras. Todos os cálculos foram feitos através do Matlab versão 6.5 (MathWorks, Natick, MA, USA).

2.3 Modelagem

Estudou-se o MCR através da construção de um modelo flexível (sem informação relativa aos espectros puros) e um modelo rígido (com importação dos espectros puros). O método adotado como referência foi o PLS, com validação cruzada (*leave-one-out*), com o qual se construiu também um modelo. Os parâmetros estatísticos (RMSEC, RMSEP e respetivos valores de r^2) resultantes da aplicação do segundo foram tidos como referência, para avaliação do primeiro.

Os modelos PLS foram construídos, com o programa de software Solo (Eigenvector Research, Inc, Wenatchee, WA). Os resultados foram estimados pelo mesmo programa.

Os modelos MCR foram construídos também com o programa Solo e os seus resultados exportados e analisados no Excel (Microsoft, Inc., Bellevue, WA).

2.4 Tratamento dos resultados

Os valores estimados pela MCR, foram trabalhados no Excel. O procedimento para a correção dos *loadings* respeita os seguintes passos, para cada componente:

1. Criar uma “célula erro”, com uma formatação equivalente à da Eq. (3). Eis um exemplo: =SQRT(SUMXMY2(C6:C5969,E6:E5969)/5964)
2. Criar uma “célula fator multiplicativo” e definir, por defeito, como 1 e uma “célula fator aditivo” e definir, por defeito, como 0
3. Definir uma coluna “S-ajustados”, de acordo com a Eq.(5):
4. Representar graficamente a coluna “S-ajustados”, cujos valores serão inicialmente iguais aos dos *loadings*, contra o espectro puro do componente em questão (valores copiados a partir da representação do respetivo ficheiro mat. no Solo: open>as raw data>plot>edit>copy plotted data).
5. Com o gráfico selecionado, utilizar a ferramenta *Solver*, definindo como objetivo a “célula erro”, que se quer minimizar, fazendo variar as células referentes aos fatores multiplicativo e aditivo. Definiu-se ainda que os valores calculados poderiam ser negativos. O *Solver* calcula novos valores para os referidos fatores e ajusta a curva inicialmente estimada pelo MCR aos valores reais.

$$S_{ajustada} = S_{estimada} * fator\ multiplicativo + fator\ aditivo \quad (5)$$

Para os *scores*, o procedimento é semelhante, com as devidas adaptações: no ponto 1. o denominador da fórmula é alterado para 40 ou 29, quer se tratem das amostras de calibração ou de teste; no ponto 3. a coluna “C-ajustadas” é definida de acordo com a Eq. (6):

$$C_{ajustada} = C_{estimada} * fator\ multiplicativo + fator\ aditivo \quad (6)$$

Foi ainda calculado o erro relativo, dividindo o erro obtido pelo valor de fração mássica típica, apresentado na **Tabela 1**.

2.5 Pré-processamento de dados

As técnicas de pré-processamento utilizadas no modelo PLS foram MCS (mean) com Mean Center para todos os componentes. O fundamento desta escolha encontra-se descrito na secção 1.2. Nos modelos RMC, foi também MSC (mean) para todos os componentes, de modo a minimizar fontes de variação dos resultados, entre as duas técnicas.

3 Resultados e discussão

3.1 Modelos MCR

Desenvolveram-se dois modelos: um modelo flexível – modelo A – que estimou os *loadings* e os *scores*, na etapa de calibração; um modelo rígido, assim designado por condicionar o cálculo dos *scores* à matriz de espectros puros introduzida no sistema - modelo B. São obtidos *scores* quer para a etapa de calibração, quer para a etapa de teste.

3.1.1 Modelo A

3.1.1.1 Calibração

Introduzida a matriz de espectros das 40 amostras de calibração no bloco X de calibração, e definidas as restrições e número de componentes do sistema, *loadings* e *scores* de cada componente foram estimados e copiados para o Excel.

Loadings

Os *loadings* estimados pelo método, para cada componente, foram ajustados aos espectros puros, por meio do procedimento descrito na **secção 2.4**. Os valores calculados pelo Solver para os fatores multiplicativo e aditivo associados ao ajuste das curvas de perfis espectrais estimados às curvas dos espectros puros são apresentados na **Tabela 2**. A **Fig.3** ilustra as curvas dos espectros puros e dos perfis espectrais estimados, depois do ajuste, para cada componente presente na formulação.

Tabela 2. Fatores multiplicativos e fatores aditivos calculados pelo *Solver*, para o ajuste dos perfis espectrais estimados aos espectros puros

MCR modelo A <i>Loadings</i>				
	Paracetamol	Celulose microcristalina	Talco	Estearato de Magnésio
Fator multiplicativo	1.68	0.85	0.10	0.36
Fator aditivo	-0.13	0.01	0.06	0.14

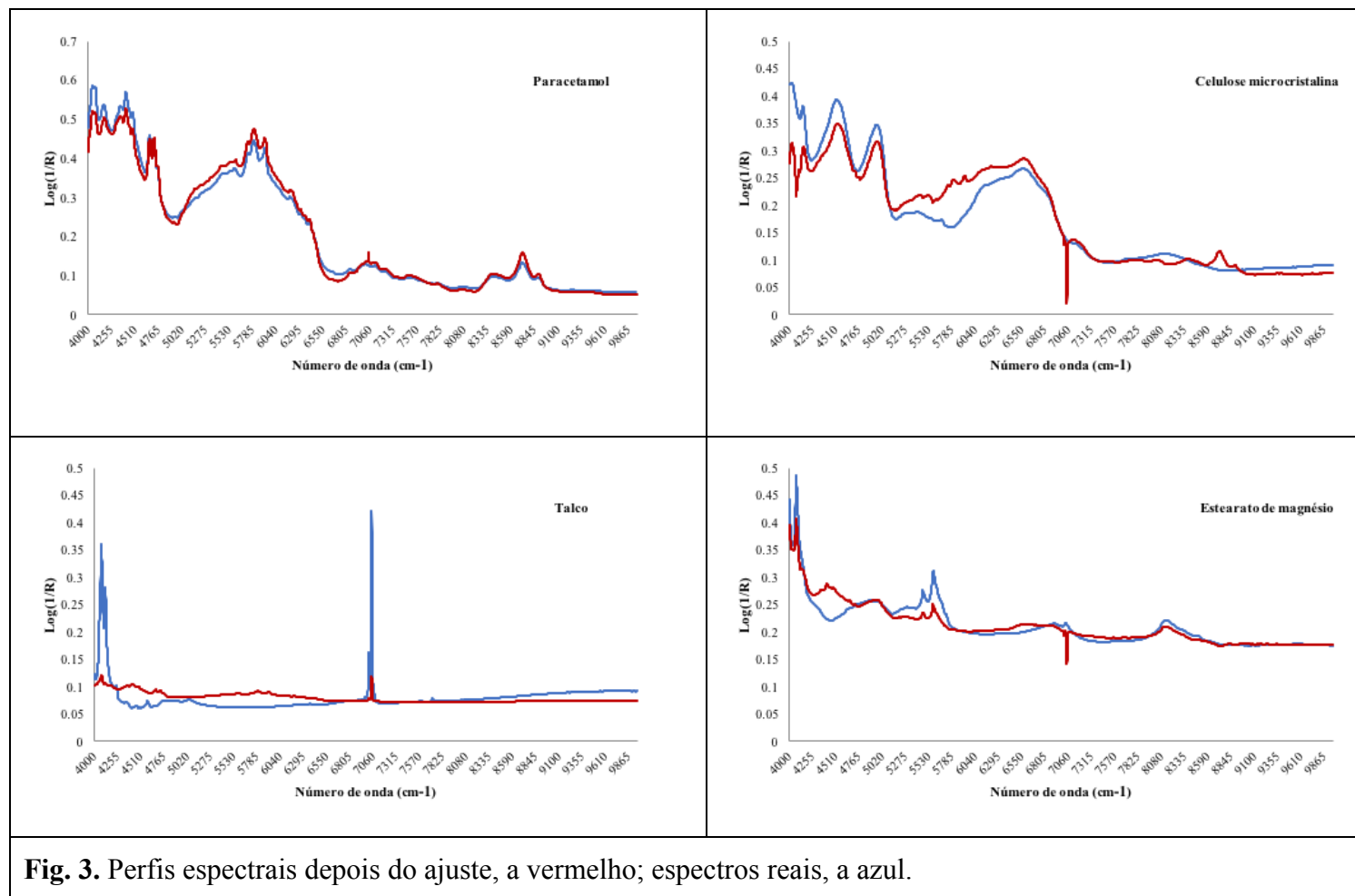


Fig. 3. Perfis espectrais depois do ajuste, a vermelho; espectros reais, a azul.

Scores

Os *scores* estimados pelo método foram ajustados aos valores de concentração conhecidos. (ver **secção 2.4**). Os valores calculados pelo *Solver* para os fatores associados ao ajuste dos perfis de concentração estimados aos perfis de concentração reais são apresentados na **Tabela 3**. A **Fig.4** ilustra as curvas dos perfis de concentrações estimado (depois do ajuste) e real.

Tabela 3. Fatores multiplicativos e fatores aditivos calculados pelo *Solver*, para o ajuste dos perfis de concentração estimados aos perfis reais, no modelo A.

MCR modelo A <i>Scores</i>				
	Paracetamol	Celulose microcristalina	Talco	Estearato de Magnésio
Fator multiplicativo	0.61	0.94	0.43	0.31
Fator aditivo	0.49	-0.05	-0.04	-0.01

3.1.1.2 Previsão

Para esta etapa, foi importada a matriz dos espectros das 29 amostras teste para o bloco X de validação. Foram estimados os *scores* deste conjunto de amostras, com base no modelo construído e copiados para o Excel. As concentrações estimadas foram corrigidas pelos mesmos fatores multiplicativo e aditivo, de acordo com a **Eq.(3)** (**secção 2.4**) e utilizando os valores presentes na **Tabela 3**.

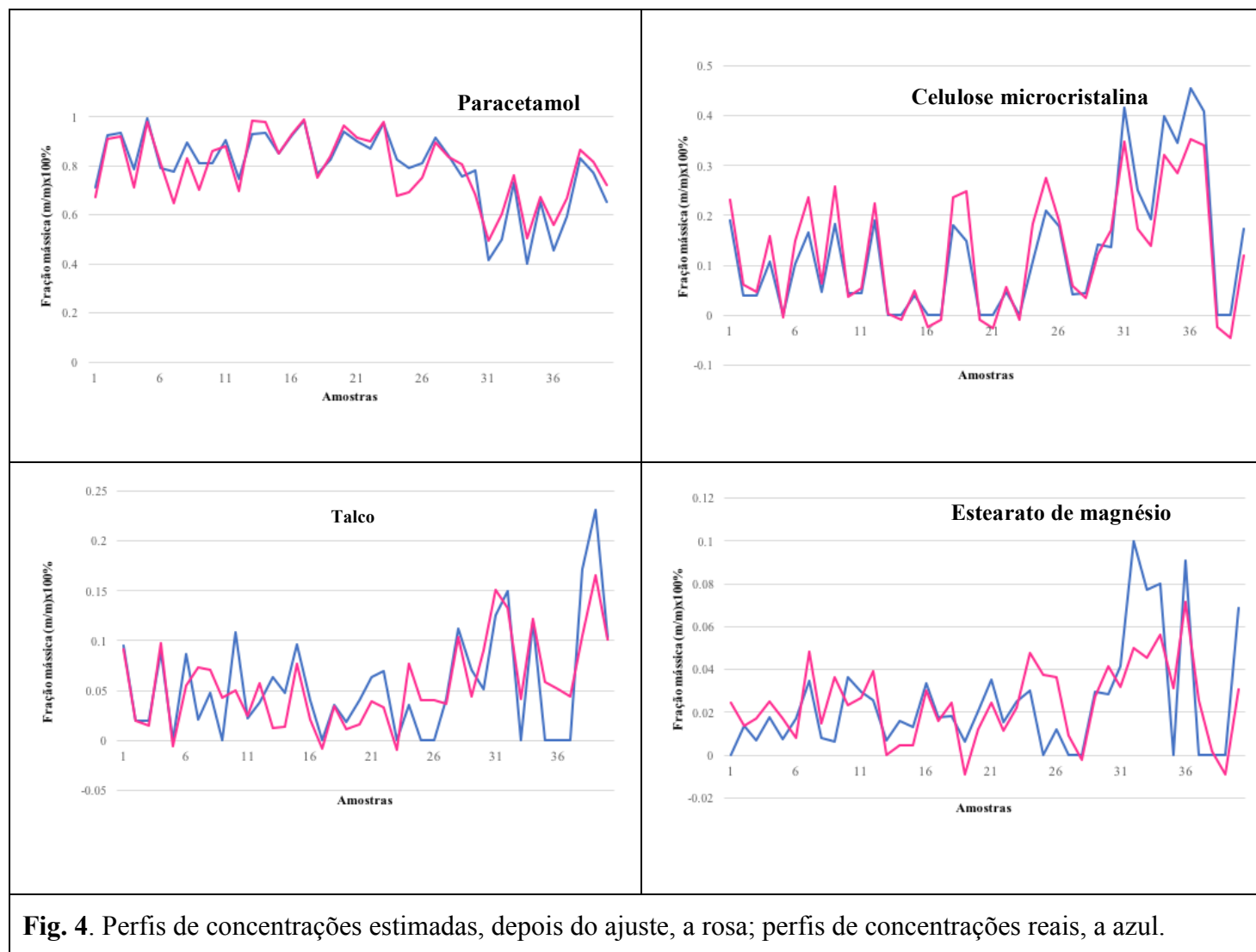


Fig. 4. Perfis de concentrações estimadas, depois do ajuste, a rosa; perfis de concentrações reais, a azul.

3.2 Modelo B

No modelo B diferiu do modelo A pela imposição da restrição de igualdade para os espectros puros. Assim, o modelo não estima a variável pura S, pois já tem esta informação disponível.

3.2.1 Calibração

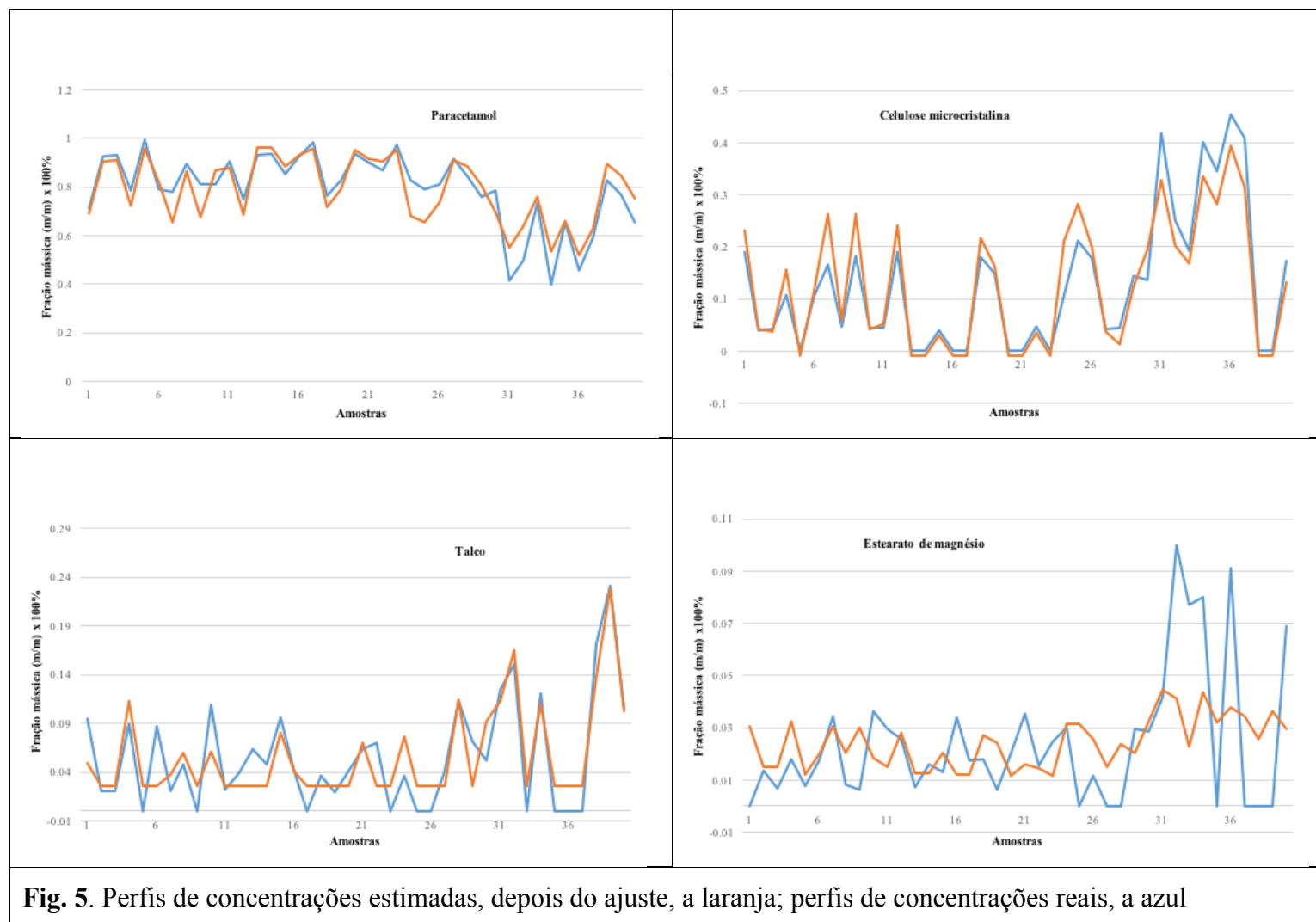
A matriz de espectros das 40 amostras de calibração foi importada para o bloco X de calibração e desenvolvido o modelo. Os *scores* estimados pelo método foram ajustados aos valores de concentração conhecidos. (ver **secção 2.4**). Os valores calculados pelo *Solver* para os fatores associados ao ajuste dos perfis de concentração estimados aos perfis de concentração reais são apresentados na **Tabela 4**. A **Fig.5** ilustra as curvas dos perfis de concentrações estimado (depois do ajuste) e real.

Tabela 4. Fatores multiplicativos e fatores aditivos calculados pelo *Solver*, para o ajuste dos perfis de concentração estimados aos perfis reais, no modelo B.

MCR modelo B Scores				
	Paracetamol	Celulose microcristalina	Talco	Estearato de Magnésio
Fator multiplicativo	1.49	0.78	0.72	-0.12
Fator aditivo	0.14	-0.01	0.03	0.07

3.2.2 Previsão

Para esta etapa, foi importada a matriz dos espectros das 29 amostras teste para o bloco X de validação. As concentrações estimadas foram corrigidas, de acordo com a **Eq. 3** (ver **secção 2.4**) e utilizando os valores presentes na **Tabela 4**.



3.3 Modelo PLS

Vários métodos de pré-processamento foram testados antes do desenvolvimento dos modelos de calibração. O método que apresentou melhores resultados de validação cruzada para cada constituinte, isto é, menor RMSECV e r^2 próximo da unidade, foi selecionado para a construção do respetivo modelo, tendo-se utilizado MSC com Mean Center.

O número de variáveis latentes a escolher no PLS, é desejavelmente o menor possível e deve ser estabelecido para um valor de RMSECV apreciavelmente baixo. O valor de RMSECV decresce à medida que aumenta o número de variáveis latentes, tendência esta que suscita dúvidas na seleção do número ótimo de variáveis latentes. Definiu-se, portanto, o número de variáveis latentes para cada modelo, de acordo com o manual de apoio ao utilizador do Solo, o qual recomenda “apenas escolher um fator adicional se o RMSECV melhorar em 2%”.(15)

Foi estudado um método de seleção espectral, na intenção de aumentar a especificidade dos modelos e por conseguinte, a sua capacidade de predição. Este consistiu em incluir, na modelagem de cada composto, apenas os intervalos de números de onda mais relevantes para cada composto, ou seja, aqueles em que estes apresentam características espectrais que contribuam para a sua identificação diferencial. Para avaliação desta hipótese, foram tidos em conta o erro e a correlação entre as matrizes e comparados àqueles originados pela modelagem feita com os intervalos espectrais completos. A conclusão foi que a seleção espectral variável não melhorou significativamente os modelos, do ponto de vista estatístico.

3.3.1 Calibração

A construção dos modelos de calibração em PLS teve início com a importação dos dados das 40 amostras de calibração para o programa de software: os espectros para a matriz X e as concentrações para a matriz Y. Foi desenvolvido um modelo para cada componente da formulação. Para este fim, selecionou-se, da matriz Y, os valores de concentração referentes a um componente de cada vez. O PLS dispõe de menus para

visualização dos *scores*, dos erros RMSEC, RMSECV e RMSEP e dos respectivos valores de r^2 , da conforme ilustra a **Fig. 6**.

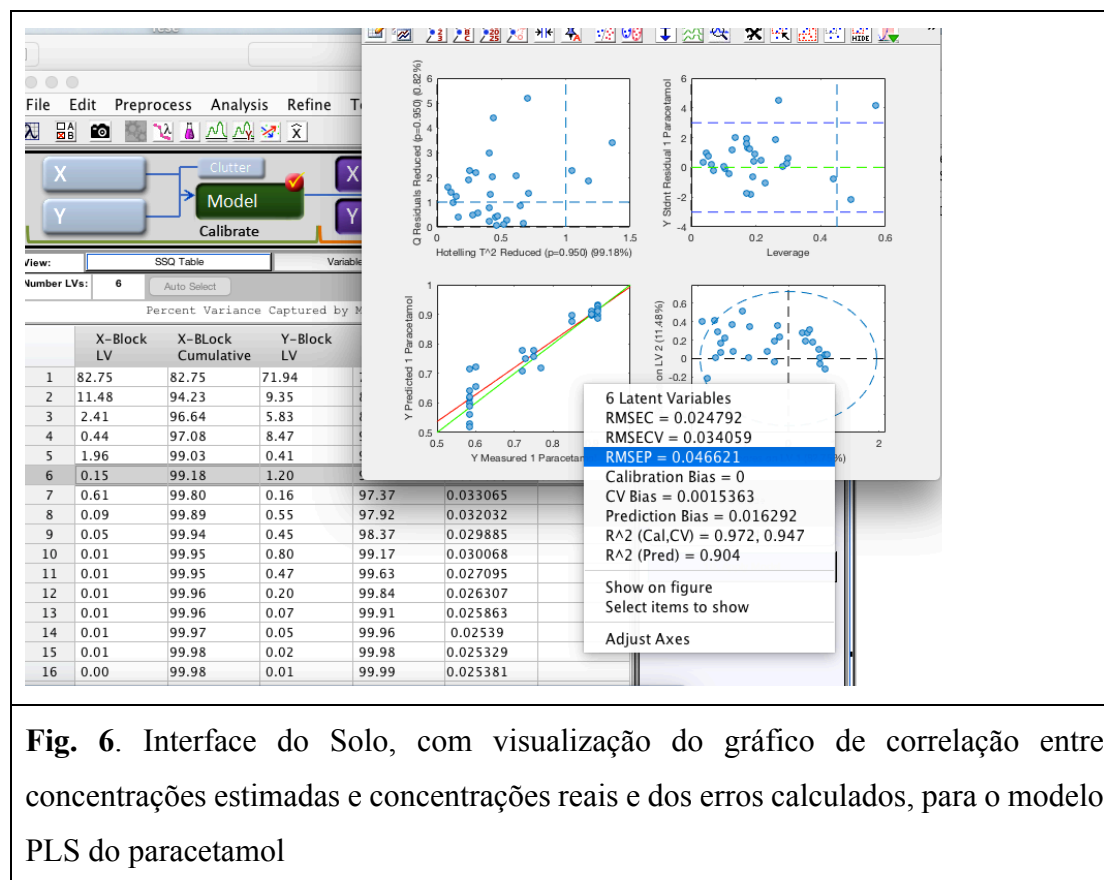


Fig. 6. Interface do Solo, com visualização do gráfico de correlação entre concentrações estimadas e concentrações reais e dos erros calculados, para o modelo PLS do paracetamol

3.3.2 Previsão

Os dados das amostras teste foram utilizados nesta etapa. Foram importados os espectros das 29 amostras, para a matriz X de validação e as respectivas concentrações para a matriz Y de validação. O próprio programa estima as concentrações das novas amostras, a partir do modelo já calibrado e compara as concentrações estimadas com as concentrações reais. A capacidade preditiva de cada modelo foi avaliada através dos valores de RMSEP.

3.4 Comparação dos métodos

Os resultados obtidos nos três modelos estão discriminados na **Tabela 5**.

Tabela 5. Resultados por modelo e por componente; 1- paracetamol, 2- celulose microcristalina, 3 – talco, 4 - estearato de magnésio

MCR modelo A					MCR modelo B				PLS			
	1	2	3	4	1	2	3	4	1	2	3	4
RMSEC	0.06	0.05	0.03	0.02	0.07	0.04	0.03	0.02	0.02	0.02	0.01	0.01
Erro relativo	7.52%	47.04%	66.45%	92.05%	8.45%	44.40%	52.42%	118.72%	2.99%	24.90%	24.09%	34.49%
R²	0.82	0.87	0.62	0.48	0.78	0.88	0.76	0.14	0.97	0.96	0.95	0.90
RMSECV									0.03	0.03	0.01	0.01
Erro relativo									4.10%	33.48%	27.94%	45.07%
R²									0.95	0.93	0.93	0.88
RMSEP	0.05	0.04	0.04	0.02	0.06	0.04	0.04	0.03	0.05	0.04	0.01	0.01
Erro relativo	6.18%	38.31%	72.28%	117.85%	6.87%	43.27%	82.83%	130.22%	5.62%	35.74%	28.34%	43.06%
R²	0.91	0.94	0.42	0.26	0.91	0.94	0.23	0.01	0.90	0.94	0.92	0.91

3.4.1 Paracetamol

A Farmacopeia Portuguesa IX define, no Ensaio A (comprimidos) de uniformidade de teor para formas unitárias, que o teor de substância ativa em cada unidade deverá ser de 85 a 115% do teor médio.(17). Este intervalo reflete um erro de 30%. Os modelos foram capazes de estimar a concentração de substância ativa, com um erro inferior àquele preconizado pela farmacopeia, sendo este de 5% no modelo A MCR, 6% no modelo B MCR e 5% no modelo PLS.

3.4.2 Excipientes

O desempenho de todos os modelos na predição das concentrações de celulose foi semelhante, obtendo-se um erro de aproximadamente 4% e um r^2 de 0.94, o que revela uma boa capacidade de predição. O mesmo não aconteceu para o talco ou para o estearato de magnésio. No PLS conseguiu-se uma boa correlação entre os dados estimados e os dados reais (valores de r^2 próximos da unidade), mas não nos modelos MCR. As concentrações estimadas também apresentaram erros muito superiores nos modelos MCR em relação ao do modelo PLS. Este facto prende-se provavelmente com a sensibilidade do sinal NIR à presença de componentes em concentrações sempre baixas. Revendo a **Tabela 1.**, a concentração do talco média é de 5% e a do estearato de magnésio de 2%.

3.4.3 Modelo A versus ModeloB

O modelo com informação dos espectros puros – modelo B – conduziu, de um modo geral a erros superiores aos do modelo A. Esta tendência contraria a expectativa inicial de que a informação adicional seria uma mais valia no desempenho do método.

4 Conclusões

O MCR revelou-se uma ferramenta útil para a monitorização das concentrações de substância ativa nas formulações testadas. As estimativas obtidas por MCR assemelham-se às obtidas com o modelo PLS e cumprem com as especificações legais de teor em formas sólidas em termos do API (substância maioritária). A estimativa das concentrações dos excipientes ficaram aquém das expectativas, em ambos os modelos, tendo sido tanto mais erróneas, quanto menor a concentração do componente na formulação. Este fato mostra alguma dificuldade da técnica MCR em prever substâncias em quantidades inferiores a 5%. Este fato tem como justificação a menor influência destas substâncias nos espectros NIR das misturas, a não-linearidade dos espectros NIR de misturas e eventualmente algum não-homogeneidade dos componentes minoritários nas misturas.

Referências Bibliográficas

1. Reich G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv Drug Deliv Rev.* 2005;57(8):1109–43.
2. Räsänen E, Sandler N. Near infrared spectroscopy in the development of solid dosage forms. *J Pharm Pharmacol* [Internet]. 2007;59(2):147–59. Available from: <http://doi.wiley.com/10.1211/jpp.59.2.0002>
3. Bosco GL, James L. Waters Symposium 2009 on near-infrared spectroscopy. *TrAC - Trends Anal Chem* [Internet]. 2010;29(3):197–208. Available from: <http://dx.doi.org/10.1016/j.trac.2010.01.003>
4. Sarraguça MC, Lopes JA. Quality control of pharmaceuticals with NIR: From lab to process line. *Vib Spectrosc.* 2009;49(2):204–10.
5. FDA. Guidance for Industry PAT: A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. FDA Off Doc. 2004;(September):16.
6. Patil AS, Pethe AM. Quality by design (QbD): A new concept for development of quality pharmaceuticals. *Int J Pharm Qual Assur.* 2013;4(2):13–9.
7. Souza AM De, Breitzkreitz MC, Filgueiras PR, José J, Rohwedder R, Poppi RJ. *Educação.* 2013;36(7):1057–65.
8. Sarraguça MC, Lopes JA. The use of net analyte signal (NAS) in near infrared spectroscopy pharmaceutical applications: Interpretability and figures of merit. *Anal Chim Acta.* 2009;642(1–2):179–85.
9. Martens H, Naes T. *Multivariate Calibration.* New York: John Wiley & Sons; 1991.
10. Jiang J-H, Ozaki Y. Self-Modeling Curve Resolution (Smcr): Principles, Techniques, and Applications. *Appl Spectrosc Rev* [Internet]. 2002;37(3):321–45. Available from: <http://www.tandfonline.com/doi/abs/10.1081/ASR-120014359>
11. Tauler R. *Multivariate Curve Resolution : theory and applications. Constraints.* 2005;

12. de Juan A, Tauler R. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. Vol. 36, Critical Reviews in Analytical Chemistry. 2006. p. 163–76.
13. Azzouz T, Tauler R. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta*. 2008;74(5):1201–10.
14. Marçó PH, Valderrama P, Alexandrino GL, Poppi RJ, Tauler R. Multivariate Curve Resolution With Alternating Least Squares: Description, Operation and Applications. *Quim Nova* [Internet]. 2014;37(9):1525–32. Available from: <http://www.gnresearch.org/doi/10.5935/0100-4042.20140205>
15. Wise BM, Gallagher NB, Windig W. Chemometrics Tutorial for PLS _ Toolbox and Solo. 2006. 420 p.
16. Rinnan Å, Berg F van den, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal Chem*. 2009;28(10):1201–22.
17. Anal T. 2. métodos analíticos.